So far:

- rank based methods
- randomization tests

to avoid a too restrictive parametric family.

Now back to parametric models:

Are there general "rules" of the properties of parameter estimates?

=> Asymptotic theory of Maximum-Likelihood estimaters

Later:

can we derive properties of the estimates from data?

=> Bootstrap

Asymptotics

Maximum-Likelihood estimators

Consider estimation problem.

Model: - density $f_{\underline{\theta}}(x)$
- probability $P_{\underline{\theta}}(X=x)$

where $\underline{\theta} = (\theta_1, \ldots, \theta_p)$ are the parameters

(e.g. $\underline{\theta} = (\mu, \sigma^2)$ for the normal distribution)

Observed data: $x_1, \ldots, x_n$ (i.i.d. observ.)
fix once observed

Likelihood: $L(\underline{\theta}) = \prod_{i=1}^{n} f_{\underline{\theta}}(x_i)$

parameter value $\longmapsto$ "likelihood of observed data under this parameter value"

ML-estimator: $\hat{\underline{\theta}} = \operatorname{argmax}_{\underline{\theta}} L(\underline{\theta})$

$= \operatorname{argmax} \ell(\underline{\theta})$

where

$$\ell(\underline{\theta}) = \log(L(\underline{\theta})) = \sum_{i=1}^{n} \log(f_{\underline{\theta}}(x_i))$$

is the log-likelihood function.

The maximum is found by setting derivatives to zero:

$$\frac{\partial}{\partial \theta_k} l(\underline{\theta}) = \sum_{i=1}^{n} \underbrace{\frac{\partial}{\partial \theta_k} \log(f_{\underline{\theta}}(x_i))}_{=: s_k(x_i, \underline{\theta})} \overset{!}{=} 0$$

"likelihood score"

$$= \sum_{i=1}^{n} s_k(x_i; \underline{\theta}) = 0$$

Remember:

For i.i.d. random variables $X_1, \dots, X_n$ with $E[X_i] = \mu$ and $Var(X_i) = \sigma^2$ we have

with any distribution

Central limit theorem (CLT)

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \approx N(\mu, \frac{\sigma^2}{n})$$

$$S_n = \sum_{i=1}^{n} X_i \approx N(n \cdot \mu, n \cdot \sigma^2)$$

for large n.

Version for ML-estimators (under some regularity conditions)

$$\hat{\theta}_k \approx N(\underbrace{\theta_k}_{\text{true value}}, \frac{\sigma^2_{asymp}}{n})$$

where

$$\sigma^2_{asymp} = \left(\int s_k(x; \theta)^2 f_{\underline{\theta}}(x)\, dx \right)^{-1}$$

$$= \Big(\underbrace{E\Big[\Big(\frac{\partial}{\partial \theta_k} \log(f_{\underline{\theta}}(X))\Big)^2\Big]}_{= I(\theta_k) \text{ : Fisher-Information}} \Big)^{-1}$$

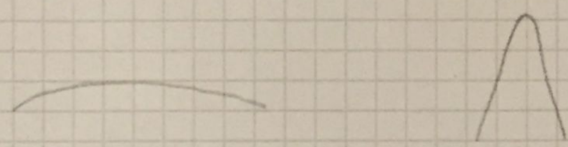
We can do (approximate) tests and confidence intervals based on that fact!

This is a very useful and general result!

<u>Remark</u>

Under regularity condition we have

$$I(\theta_k) = -E\left[\frac{\partial^2}{\partial \theta_k^2}\left(\log f_{\underline{\theta}}(X)\right)\right]$$

"Picture": 

Curvature of likelihood = estimation accuracy!

<u>Example</u>

$$f_\theta(x) = \theta x^{\theta-1}, \quad 0 < x < 1; \quad \theta > 0$$

$$L(\theta) = \prod_{i=1}^{n} \theta x_i^{\theta-1}$$

$$\ell(\theta) = \sum_{i=1}^{n}\left(\log(\theta) + (\theta-1)\cdot\log(x_i)\right)$$

$$\frac{\partial \ell}{\partial \theta} = \sum_{i=1}^{n}\left(\frac{1}{\theta} + \log(x_i)\right) \overset{!}{=} 0$$

$$\Rightarrow \hat{\theta} = -\frac{n}{\sum_{i=1}^{n}\log(x_i)}$$

Fisher-Information:

$$\frac{\partial^2 \log f}{\partial \theta^2} = -\frac{1}{\theta^2}$$

$$\Rightarrow I(\theta) = -E\left[\frac{\partial^2}{\partial \theta^2}\left(\log f_\theta(X)\right)\right] = \frac{1}{\theta^2}$$

For observed data $x_1, \dots, x_{50}$ with

$$\frac{1}{n} \sum_{i=1}^{n} \log(x_i) = -0.176$$

we have

$$\hat{\theta} = 5.69$$

The (asymptotic) variance is given by $\frac{1}{n} \cdot \theta^2$.

Hence, an approximate 95% confidence interval is given by

$$\hat{\theta} \pm 1.96 \sqrt{\frac{\hat{\theta}^2}{n}} = [4.12, 7.27]$$

(see also R-file)